

A Hybrid Technique to Classify Trending Topic on Twitter Dataset

Pramod S Nair¹, Nidhi Choubey², D Srinivasa Rao³

¹Prof. of CSE Department, MITM, Indore, India

²Research Scholar, MITM, Indore, India

³Assoc.Prof. of CSE Department, MITM, Indore, India

¹pramodsnair@yahoo.com

²nidhichoubey.cs@gmail.com

³sridas712@gmail.com

Abstract— Text mining is in trend of research and development these days. A number of recent contributions in this domain are observed. Different social media text analysis techniques are used to find the trending topic. Both kinds of data mining techniques (i.e. supervised and unsupervised) are used in text mining. The proposed study is focused on the supervised learning approach. In the first step the dataset is prepared, thus different subjects or domain based data is extracted from twitter and according to these subjects the class labels are appended with the data. Further the dataset is split into two sets i.e. the training data and the testing data. In the next step the data is pre-processed where stop words and special characters are removed from the dataset. This dataset is used with the three different classification algorithms KNN (k-nearest neighbor), SVM (support vector machine) and a combination of both SVM and KNN. Hybrid approach is applied to build the classification model and it is implemented on test dataset to discover the twitter subjects. The result of the proposed work is compared with traditional KNN and SVM algorithm. According to the results obtained, the proposed hybrid technique provides more precise and better results as compared to other conventional classifiers.

Keyword - SVM, Sentiment Analysis, Twitter, Social Media, Text Mining, Chat Classification, k-NN

I. INTRODUCTION

Social Media is one of the major sources of information. People are dependent on social networking sites to interact with other users and to exchange their views regarding any subject. Facebook chats, Twitter tweets are the examples from where we can judge the mentality and behaviour of any individual. Text mining and Sentiment mining is in trend of research and development. A number of recent contributions in this domain are observed. Among them the popular research works are relevant to the social media text analysis. In this presented work the social media text analysis techniques are investigated to find the trending topics on twitter. This task is performed with the help of data mining techniques. Both kinds of data mining techniques (i.e. supervised and unsupervised) are supportable in text mining. But the proposed study is focused on the supervised learning approach. In the first step the dataset is prepared, thus the different subjects or domain based data is extracted using the twitter and according to their subjects the class labels are appended with the data. Further, the dataset is split into two sets of data, training and testing datasets. Next, the data is transformed to make it available for mining purpose, thus the pre-processing techniques are employed to remove the stop words and special characters from the dataset. After refinement of data the dataset is converted into an encoded format. Similar process is applied on test dataset to create suitable training and testing data. This dataset is used with the three different classification algorithms KNN (k-nearest neighbour), SVM (support vector machine) and a combination of both SVM and KNN. After implementation of all the techniques the classification task is performed on testing dataset to discover the twitter subjects.

Language analysis for online networking can be done in better way using computer systems instead of by humans. Thus manual strategy is also quite costly and takes additional time. This objective is fulfilled by content mining. So there are many researched which are done on content mining [2]. In this segment, diverse content digging advancements are talked about for mining content.

A. Information Extraction

Information Extraction is the basic and first step of data mining. It means to extract structured data from unstructured data and huge amount of data which may be noisy or in missing condition. It is one of the most important methods also because this extracted data is obtained from various sources and is used for applying mining techniques. The primary goal of information extraction is to find out the patterns from the huge amount of text data and also determines the relationship among the data items [3].

B. Summarization

This procedure has primary objective of exact content from substantial number of content records. Physically it is unrealistic to compress vast archives [4]. In many research focuses, it is impractical to peruse all content reports that mean specialist has no opportunity to peruse these documents. They compress records and makes synopsis of records from principle points.

C. Topic Tracking

Topic tracking refers to keeping track of user's previously made searches in order to find out the search pattern of the user. The previously searched topics of any individual are kept in a file known as user profile. This procedure is used to track the future searches that an individual can need. One of its drawbacks is that it also returns unnecessary data too. Theme following is utilized as a part of numerous territories, for example, radio, news communicates and so forth.

D. Classification

It is a procedure of discovering primary subject of record by including metadata and investigating report. This system finds out the total number of words present in a document and from that number, it chooses the subject of the text. This method arranges text files into predefines class labels [5]

The remaining document is organized as Related Study is given in the II chapter followed by proposed work, methodology and the then Result is discussed. At last conclusion and future work is given.

II. LITERATURE SURVEY

The variety of ways that the issue has been contemplated and the numerous systems that have been created makes correlation between these reviews testing.

Literature reveals that the likelihood of anticipating a few clients and message traits in content based, constant, web based informing administrations. For this reason, a substantial accumulation of chat messages are inspected. The materialness of different directed grouping methods for extricating data of the tweets is accessed. A term-based approach is utilized to research the client and message properties with regards to vocabulary utilize while a style-based approach is utilized to analyse the tweets indicated by the varieties in the writers' composition styles. From 100 of creators, the character of a creator is effectively anticipated with 99.7% exactness [6].

“Text classification methodologies applied to micro-text in military chat” includes strategies to group lines of military talk, or posts, which contain things of intrigue. Creators assessed a few current content classification and highlight determination strategies on visit posts. This talk posts are cases of 'miniaturized scale content', or content that is by and large short long, semi-organized, and described by unstructured or casual linguistic use and dialect. Despite the fact that this review concentrated particularly on strategic updates by means of talk, they trust the discoveries are appropriate to substance of a comparative semantic structure. Fulfilment of this point of reference is a huge initial phase in considering more perplexing arrangement and data extraction [7].

Sentiment Classification goes for digging surveys of individuals for a specific occasion's subject or item via programmed characterizing the audits into good or bad sentiments. Automatic opinion mining is beneficial for chief and customary individuals. There are principally two sorts of methodologies for slant characterization, machine learning techniques and semantic introduction strategies. In spite of the fact that some pioneer investigates investigated the methodologies for English surveys order, few occupations have been done on assumption characterization for Chinese audits [8].

This study considers the issue of consequently ordering human slant from normal dialect composed content. In this assessment mining space, they think about the exactness of gathering models, which exploit gatherings of learners to yield more noteworthy execution. They demonstrate that these troupe machine learning models can fundamentally enhance slant grouping for freestyle content [9].

Micro blogging today has turned into an extremely prevalent specialized device among Internet clients. A huge number of clients impart insights on various parts of life consistently. Accordingly smaller scale blogging sites are rich wellsprings of information for conclusion mining and slant examination. Since smaller scale blogging has showed up generally as of late, there are a couple of researchers that are dedicated to this subject.

This study is concentrating on utilizing Twitter, the most well known miniaturized scale blogging stage, for the errand of Emotion examination. They will demonstrate to naturally gather a corpus for Emotion examination and sentiment mining purposes and after that perform phonetic investigation of the gathered corpus and clarify found wonders. Utilizing the corpus, writer will fabricate an Emotion classifier that will have the capacity to decide the feeling class of the individual written work [10].

III. PROPOSED WORK

Text mining and its approaches are used in various applications for extracting the essential information from the text data. This work is dedicated to obtain the topic classification on the social media data. That helps to identify the trending topics on twitter text. This section involves the understanding of the proposed technique for mining sentiment based topic tracking.

A. System Overview

Data mining is performed on any kind of data for automated data analysis. In this analysis different kinds of algorithms are involved to process the data and to obtain the similar pattern of data. According to the requirements of the application the selection of an algorithm is employed. The algorithm may be supervised in nature or unsupervised; the supervised algorithms are employed when some patterns of data are predefined and need to distinguish the patterns among pre-known classes. On the other hand the unsupervised technique is used when the data is not initially available on predefined classes and the data is in huge volumes. In addition of that the supervised techniques are used for accurate recognition of data.

In the presented work the supervised classification technique is used for analysis and text pattern classification. The main aim of this text classification technique is to discover the trending topics in the social media according to their hidden sentiments. This approach also helps to distinguish the patterns of data, in other words the topic information in any tweet. In order to perform this text classification the pre-processing technique is optimized in a different manner. Additionally the three different kinds of classification models are used. In first the SVM (support vector machine), in next the KNN (k-nearest neighbor) classifier is used for performing the classification task. It is a hybrid approach of SVM and KNN which is used to improve the precision and other parameters of classification.

B. Methodology

For The method applied to the proposed system is described in this section. Figure 2.1 shows the proposed technique of data processing and classification of text.

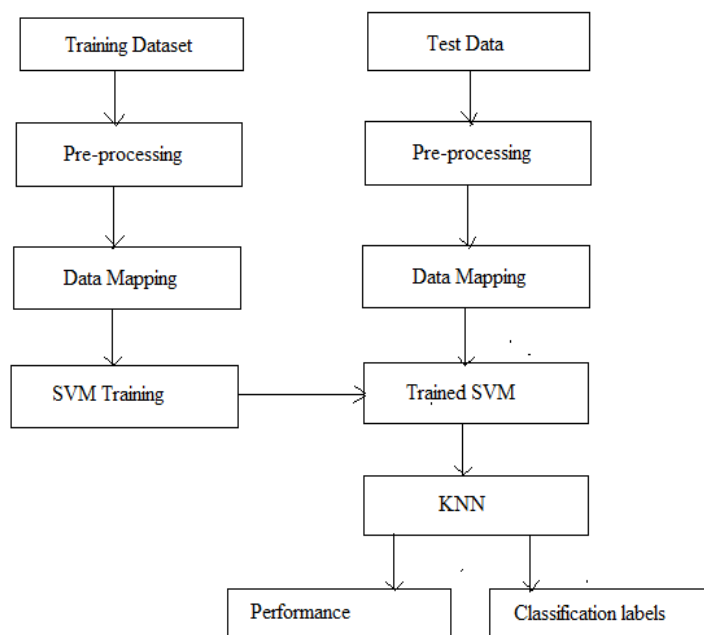


Fig.1. Proposed Data Model

These components are organized in such manner by which the outcome of one component becomes input for next process.

1) *Training Dataset*: In order to demonstrate the capability of the proposed technique for topic tracking the twitter dataset is used. Before classifying this data using the proposed system, different topics of the data are identified and the classes of data are labelled for each tweet. This data is further used for classification purpose.

2) *Test data*: The complete dataset is divided into training and testing data in such a way that the results are not biased. This test data is used for performing the evaluation of data model for finding the accuracy of training.

3) *Pre-processing*: Pre-processing is task of data transformation and correction. Therefore the different techniques of data cleaning and refinement are applied to data for finding the data which are acceptable with the algorithm used. In order to process the data, two different techniques are applied in this phase.

4) *Removing the special character*: This is the first phase of data pre-processing. Here a list of special characters is prepared, which is used with the data to find and replace functionality for removing the unwanted special characters.

5) *Removal of stop words*: After removal of unnecessary special characters the stop words are reduced from the text data. The stop words are those words that are frequently used during the sentence formation. But these words do not put impact on recognizing the domain, i.e. this, that is, are, who, and so on. Here a list is prepared with the considered stop words and by using “find and replace” method all the unwanted stop words are removed.

6) *Data mapping*: After pre-processing of data, it is produced as input to the data mapping phase. Here data is again transformed into a different format. Thus we need to prepare the mapping of the text contents into symbolic form. In order to understand the process of mapping we take an example

TABLE I. Example of Mapping

Actual sentence	Mapped sentences
1. hello i m online	1. “1”, “2”, “3”, “4”
2. i m enjoying online	2. “2”, “3”, “5”, “4”

Table I gives that the individual words are converted into a definite symbol and during repetition of words the same mapping symbol is repeated. For example word “online”, repeated in both the sentences at fourth place, thus both of the places are replaced with number “4”.

7) *SVM training*: The encoded strings of the same length with the class label symbol are provided as input training dataset for SVM training. After some training cycles when the maximum iterations are completed, then, we understand the training process is completed. After training of SVM, the machine is ready to perform classification on similar pattern data.

8) *Trained SVM*: The trained SVM classifier is now used with the testing dataset. But the similar tweeted texts are selected to prepare the testing dataset. Thus to create a similar kind of input patterns (i.e. as similar to training inputs for SVM classifier) the testing data is also converted to the similar encoded text. In this phase the encoded test data is produced for generating the class labels for input tweets. These class labels are name of those domains which are needed to find out. To demonstrate to classified outcomes and further process we take an example.

TABLE II. SVM Outcomes

Input to SVM	Predicted outcome	Label actual name
1. “2”, “3”, “5”, “4”	A	Politics
2. “1”, “3”, “2”, “4”	A	Politics
3. “2”, “6”, “4”, “1”	B	Movies
4. “2”, “3”, “3”, “4”	B	Movies

According to Table II, the first column of table provides the input sequences to classify in SVM. After computing the class labels the data is predicted class label is given in second column. Finally the actual class labels are obtained by prediction outcome. But from this data only the initial two columns are used for further outcome refinement.

9) *KNN*: In this phase the outcome of previous phase data is accepted. This process helps to sub-divide each class into two sub-classes. First of them is called more likely and second is called less likely. Therefore similar class label patterns are used to find the distance between each other. Here to differentiate between two groups of patterns a distance threshold 75% is taken.

10) *Performance*: After classification of data the performance of classification is measured under the different performance parameters.

C. Proposed Algorithm

The designed process of classification is summarized in terms of step process involved for training and testing. Table III gives the steps of proposed algorithm.

TABLE III. Proposed Algorithm

Input: training dataset D_{train} , test dataset D_{Test} , Threshold Th Output: class labels C
Process: $T_{input} = readDataset(D_{train})$ $D_{preprocessed} = preprocessData(T_{input})$ $D_{Train} = EncodeDataset(D_{preprocessed})$ $Model_{Trained} = SVM.Train(D_{train})$ $[P, Cin] = Model_{Trained}.classify(D_{Test})$ $C = KNN.classify(P, Cin, Th)$ Return C

IV. RESULT ANALYSIS

This part gives the detailed results and comparison of the proposed User Chat Classification of Social Networks. Different performance parameters are calculated and compared for the proposed algorithm and traditional algorithms.

A. Precision

Precision measure is the ratio of the number of correct positive results and number of all positive results. It measures the exactness of any classifier. The higher the precision means that less false positives (FP), whereas the lower precision means that more the false positives are. Here, we show precision rate formula.

$$PrecisionRate = \frac{TP}{TP + FP}$$

-TP is the number of true positives

-FP is the number of false positives

The evaluation of text classification is demonstrated in Fig 2 and Table IV. In this Figure, X-axis shows the number of different runs during the algorithm process and Y-axis depicts the precision rate of all three classifiers. Blue line show proposed hybrid approach using SVM and KNN, green line show the traditional KNN classifier performance and red line shows base classifier of SVM.

TABLE IV. Tabular Values of Precision Rate

Number of Runs	SVM + KNN	SVM	KNN
1	0.89999	0.5125	0.8387
2	0.95517	0.47777	0.80769
3	0.91538	0.43529	0.83333
4	0.88368	0.44895	0.85698
5	0.92458	0.49851	0.82971
6	0.92986	0.48796	0.83146

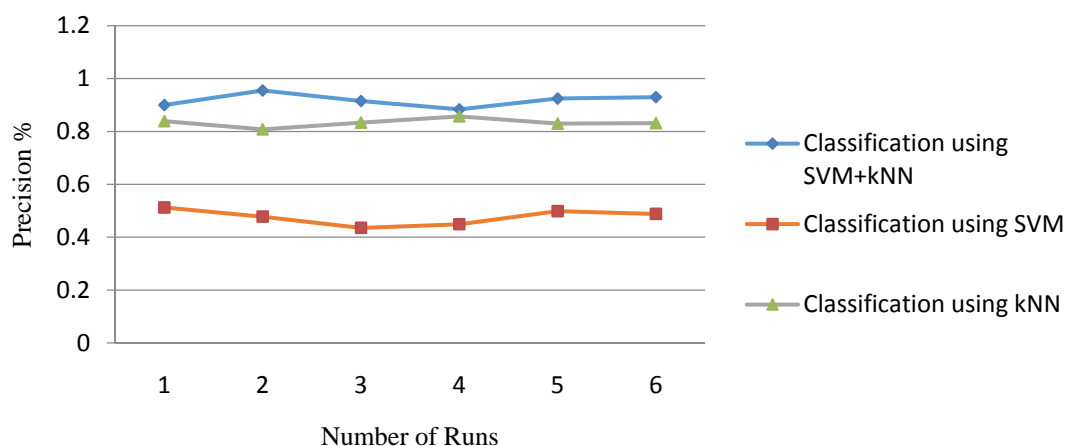


Fig. 2 Graph for Precision Rate

In this we have compared and analysis their comparison on the basis of the produced output. By this result, the sentiment of the twitter data set of user chat data, the proposed hybrid classifier performance is most efficient and accurate in terms of the precision rate other than evaluated classifier.

B. Recall

Recall is the ratio of the number of correct positive results and number of positive results that should have been returned. It measures the completeness or the sensitivity of the sentiment classifier. Higher the recall means that small false negatives (FN), whereas lower the recall means more false negatives. In this, following recall rate formula used to calculate classification of social chat using SVM and KNN:

$$\text{RecallRate} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

-TP is the number of true positives

-FN is the number of false negatives

TABLE V. Tabular Values of Recall Rate

Number of Runs	SVM + KNN	SVM	KNN
1	0.84375	0.62548	0.68421
2	0.7037	0.61514	0.66636
3	0.70588	0.60872	0.66523
4	0.72564	0.60112	0.67698
5	0.75982	0.58741	0.63547
6	0.76954	0.61578	0.67035

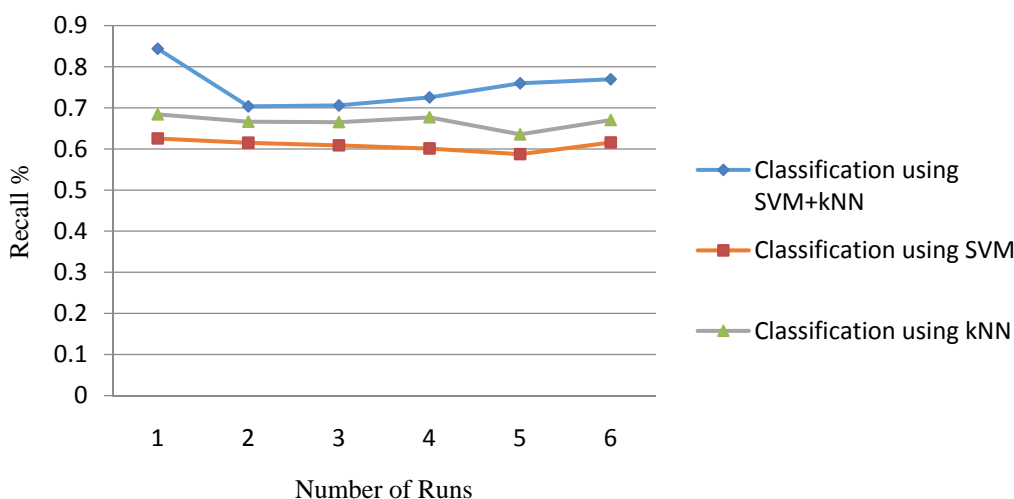


Fig. 3 Graph for Recall Rate

Figure 3 and Table V shows the comparative recall rate of implemented classification algorithm. The X-axis contains the different execution of the project for training performance and the Y axis shows the performance in terms of recall rate percentage. The recall rate of the traditional classifier is given by the green line and the performance of the proposed classification Technique is given using the blue line and the base classification show by red line. The performance of the proposed classification is better and efficient and reduces with the amount of data increases.

C. F-Measure

F-Measure or F1 Score is the harmonic mean of the Precision and Recall often used as a weighted average for balancing quality vs. quantity of true positives selection of an algorithm given by:

$$F_{\text{Measure}} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

TABLE VI. Tabular Values of F-Measure

Number of Runs	SVM + KNN	SVM	KNN
1	0.87096	0.58603	0.75362
2	0.81037	0.54578	0.71186
3	0.79709	0.50285	0.69444
4	0.77598	0.52471	0.72854
5	0.81369	0.55876	0.70827
6	0.88124	0.57634	0.72545

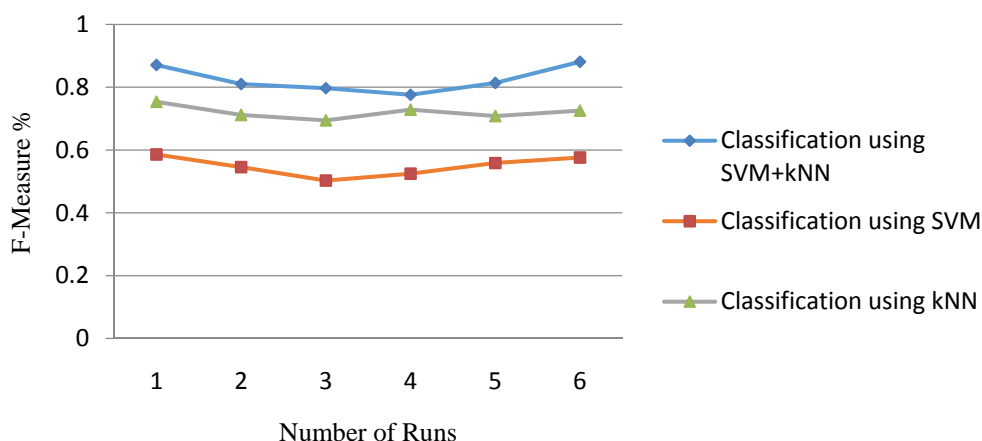


Fig. 4 Graph for F-Measure

Figure 4 and Table VI shows the performance of Proposed Hybrid text Classification systems in terms of F-measures. To demonstrate the performance of the system the X- axis shows the total runs for data execution and the Y axis represents the performance in terms of F-measures. According to the results we obtain, the performance of the implemented system is much stable and enhancing approach of user chat text classification lies in the combination of SVM and KNN classifier. In addition of that the results are in a more progressive manner as the amount of database is increasing. Thus the obtained results are adoptable and efficient for the text data sentiment analysis.

D. Training Time

Training is the process of taking content that is known to belong to specified classes and creating a classifier on the basis of that known content. Training time defines the time which is used to train the classifier to process twitter dataset. This time is used to train different size data along with classifying and processing of the data execution.

$$\text{TrainTime} = \text{ClassifierEndTime} - \text{ClassifierStartTime}$$

The Train time of the implemented algorithm is given by Figure 5 and Table VII. X-axis shows different execution of system and Y- axis contains train time utilization in milliseconds. According to the results, the time required by the proposed system to train the dataset is less compared to other traditional algorithms. The amount of time increases in similar manner as the amount of data for analysis increases.

TABLE VII. Tabular Values of Training Time

Number of Runs	SVM + KNN	SVM	KNN
1	314725	694269	289265
2	366129	666558	289265
3	434135	711141	289263
4	458752	736528	289570
5	501489	703254	289575
6	487521	758413	289280

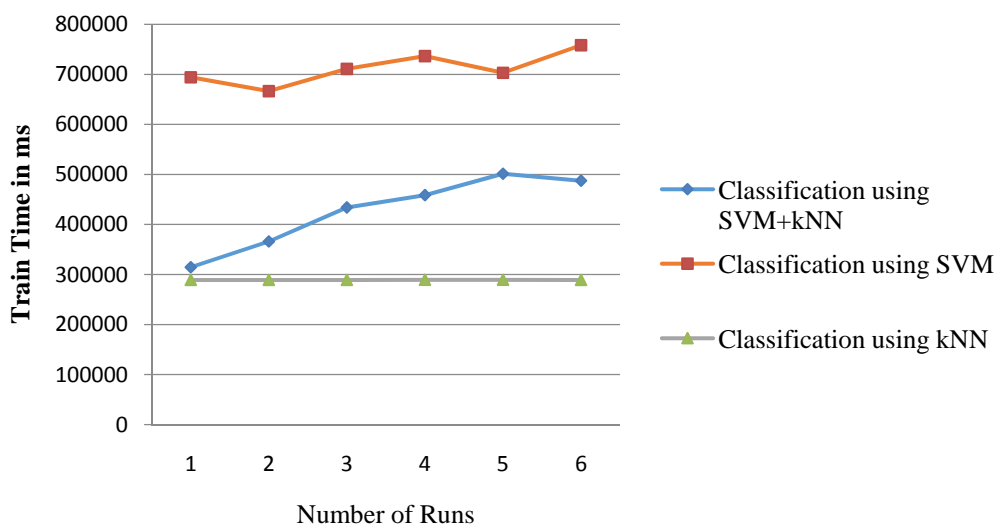


Fig. 5 Graph for Training Time

E. Testing Time

In data mining, every system requires an evaluation to produce the accurate result which is adoptable by the end user. Similarly, in this scenario test time is the time consumption where test the classified trained generated by training process. Test time show the correctness of the evaluated data and their output. Test time can be calculated by following formula:

$$\text{TestTime} = \text{ClassifiedTraineddataEndTime} - \text{ClassifiedTraineddataStartTime}$$

TABLE VIII. Tabular Values of Testing Time

Number of Runs	SVM + KNN	SVM	KNN
1	309	303	358
2	310	306	352
3	335	330	369
4	336	333	375
5	335	340	380
6	340	349	383

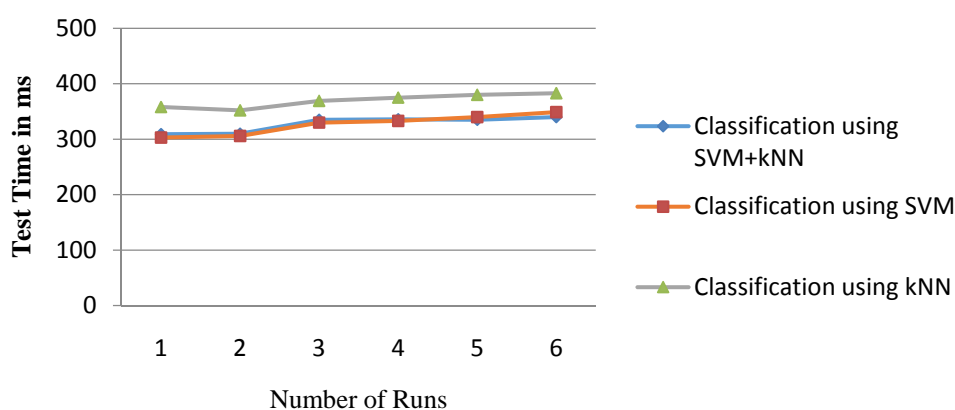


Fig. 6 Graph for Test Time

Testing time is shown using Figure 6 and Table VIII. In this figure, X-axis shows the number of different runs and Y-axis shows the different experiments test time of all three classifier. Blue line show proposed SVM and KNN combine approach, green line show the conventional KNN classifier performance and red line show base classifier of SVM. In this we have compared and analysed the comparison on the basis of the produced output. In this performance, proposed approach and base classifier taking approximately similar time to test the data and traditional KNN consuming more time. So that combine approach minimizes test time as compared to both performances.

F. Memory Consumption

Memory consumption is the amount of memory required by the system to execute properly. Sometimes it is also given by the space complexity. The formula is given below:

$$\text{MemoryConsumption} = \text{TotalMemory} - \text{FreeMemory}$$

TABLE IX. Memory Consumption

Number of Runs	SVM + KNN	SVM	KNN
1	20965	21458	22514
2	21562	23845	21658
3	17285	19031	18003
4	18547	20154	19575
5	16521	20144	18524
6	19732	25356	22162

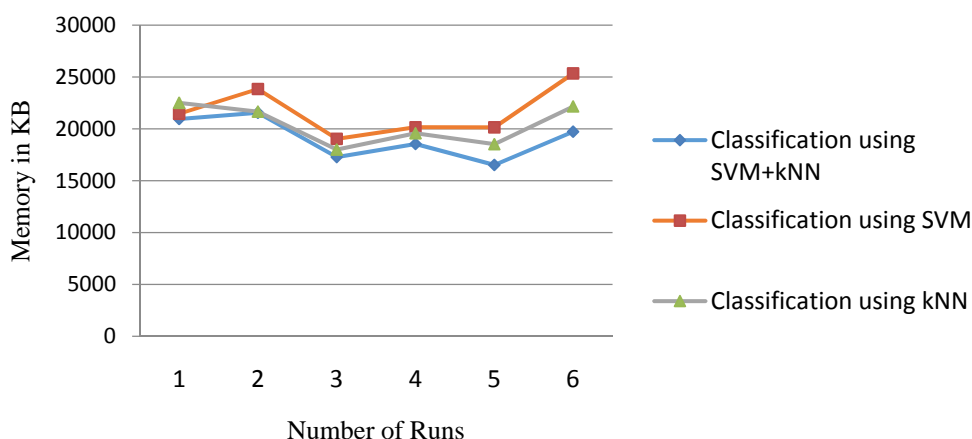


Fig 7 Memory Consumption

Memory utilization is the amount of information stored in the main memory, and thus it affects the cost of computation. The performance of our proposed classifier is given by Figure 7 and Table IX. In this figure, X-Axis displays total number of runs by executing algorithms and Y-Axis shows the memory utilization in kilobytes (KB). As indicated by the acquired outcomes the execution of calculation shows comparable conduct with various framework execution, yet the measure of memory utilization is increments with the measure of information.

V. CONCLUSION

Online social networking groups mostly display complex aggregate behaviour. Since feelings assume a relevant part in human basic leadership, social media plays a big role in driving human interest from one subject to another. A standout amongst the most applicable undertakings in Sentiment Analysis is Polarity Classification, went for grouping the opinion behind writings. Interpersonal organizations and small scale blogging apparatuses, for example, Twitter permit people to express their suppositions, sentiments, and considerations on an assortment of themes as short instant messages. These short messages (ordinarily known as tweets) may likewise incorporate the passionate conditions of people, (for example, satisfaction, nervousness, and sorrow, happiness) and also the feelings of a bigger gathering.

The proposed Hybrid Social Chat Classification Approach using SVM and KNN uses the best features of both the traditional classifiers and analyses the social network based text for their sentiments and orientation based text classification. It involves all the data mining stages i.e. pre-processing; Feature Extraction, tagging, learning from training dataset, creating classification model and the classification of newly arrived patterns. Pre-processing involves removal of punctuations and unnecessary frequent words which are not important for deciding the class of the tweet, tagging involves the conversion of the data in terms of the sentence part of speech and feature selection from the raw data. After selecting appropriate feature, three different data mining models are applied to data that are used for learning on the trained patterns and then using this model to classify the new upcoming patterns.

The performance of the system is estimated for finding the parameter, i.e. Precision, Recall, F-Measure, Train Time, Test Time and Memory of classification algorithms. The performance comparison of system with other traditional methods is given using Table X.

S. No.	Parameters	Hybrid SVM+KNN Approach	Based SVM Approach	Traditional KNN Approach
1	Precision	High	Low	Low
2	Re-call	Low	Low	High
5	F-Measure	High	Low	Average
3	Train Time	Average	High	Low
4	Test Time	Average	Average	High
5	Memory Consumption	Low	High	Average

The proposed model for text classification is desirable and efficient for classifying the text data from various domains and finding out which domain or subject is more popular among users. The proposed work makes empirical contributions to this research area by comparing the performance of different popular sentiment based user chat classification approaches and developing a collaborative approach, which further improves the sentiment classification performance. The overall performance of the proposed methodology is satisfactory and better compared to the traditional classification algorithms

VI. FUTURE WORK

The given proposed work is effective for characterizing the classes of content information for examining framework execution. This way the proposed strategy includes various methods for enhancing the classification rate. In near future,

- The proposed work can be elaborated as a web administration service, which if asked can access various networking websites for gathering surveys and synopsis and display to end user.
- A future study will also concentrate on the utilization of the certain conclusion examination inside the sentiment analysis in informal organizations.

REFERENCES

- [1] Eman M.G. Younis, "Sentiment Analysis and Text Mining for Social Media Micro-blogs using Open Source Tools: An Empirical Study", International Journal of Computer Applications (IJCA), Volume 112 – No. 5, February 2015.
- [2] Mrs.K. Mythili, and Mrs. K. Yasodha, "A Pattern Taxonomy Model with New Pattern Discovery Model for Text Mining", International Journal of Science and Applied Information Technology, Volume 1, No.3, July – August 2012.
- [3] N. Kanya and S. Geetha, "Information Extraction: A Text Mining Approach", IET-UK International Conference on Information and Comm. Technology in Electrical Sciences, IEEE, Dr. M.G.R. University, Chennai, Tamil Nadu, India, 1111- 1118, 2007.
- [4] Falguni N. Patel, Neha R. Soni, "Text mining: A Brief survey", International Journal of Advanced Computer Research, Volume-2 Number-4 Issue-6 December- 2012.
- [5] Vishal gupta and Gurpreet S. Lehal, "A Survey of Text Mining Techniques and Applications", Journal of Emerging Technologies in Web Intelligence, PP. 60- 76, 2009
- [6] Rosa, Kevin Dela, and Jeffrey Ellen. "Text classification methodologies applied to micro-text in military chat", Machine Learning and Applications, 2009, ICMLA'09 International Conference on, IEEE, 2009.
- [7] ChangliZhang, WanliZuo, Tao Peng, and Fengling He, "Sentiment classification for chinese reviews using machine learning methods based on string kernel", In Proceedings of the 2008 Third International Conference on Convergence and Hybrid Information Technology -Volume 02, pages 909–914, 2008.
- [8] Whitehead, Matthew, and Lary Yaeger. "Sentiment mining using ensemble classification models", Innovations and advances in computer sciences and engineering, Springer Netherlands, PP. 509-514, 2014
- [9] R C Balabantara, Mudasir Mohammad, Nibha Sharma Multi-Class Twitter Emotion Classification: A New Approach, Published in International Journal of Applied Information Systems (IJ AIS) – Foundation of Computer Science FCS, New York, USA Volume 4– No.1, 2012.
- [10] Alec Go, RichaBhayani, and Lei Huang, "Twitter sentiment classification using distant supervision", PP. 1 – 6, 2009

AUTHOR PROFILE

Pramod S Nair received B.Tech, M.Tech and Ph.D degree in Computer Science. His research interests include Data Mining, Web Mining Internet applications, Information Retrieval and Knowledge Discovery. He has published many research papers in international conferences and journals. He is also associated with review panels of different conferences and journal of international repute.

Nidhi Choubey received BE degree in Computer Science and engineering and is currently working towards the ME degree in the Department of Computer Science at Medicaps Institute of Technology and Management, Indore Madhya Pradesh. Her Research interest includes data mining, text mining, Social Media analytics.

D Srinivasa Rao M.Tech, (Ph.D) is working as an Associate Professor in the Department of Computer Science & Engineering at Medicaps Institute of Technology and Management, Indore, Madhya Pradesh, India. He has 20 years of teaching experience. His area of interest in Adhoc Networks, Distributed Systems, Network Security, Image Processing & Data Mining. He has guided more than 60 Post Graduate Students. He has published 2 books and 25 papers in international journals. He presented 2 papers in National Conferences, 1 paper in International Conference and has attended 35 National Workshops / FDP / Seminars etc. He is a life member of Professional Society like ISTE.